Applications Status

EGEE3 conference, Athens

April 18, 2005

Johan Montagnat

(on behalf of NA4)

- Overview of application areas
- Organisational matters
- Use of the LCG service and some first formal user feedback
- gLite usage and testing
- Conclusions and forward planning

Applications deployed on EGEE

- **Three application groups**
  - High Energy Physics pilots
  - Biomedical application pilots
  - Generic applications (catch-all)

- **Multiple infrastructures, two middlewares**
  - EGEE LCG2 production infrastructure
  - GILDA LCG2/gLite integration infrastructure
  - gLite testbeds (development/testing/certification)

- **Many users**
  - broad range of needs
  - different communities with different background and internal organization

High Energy Physics

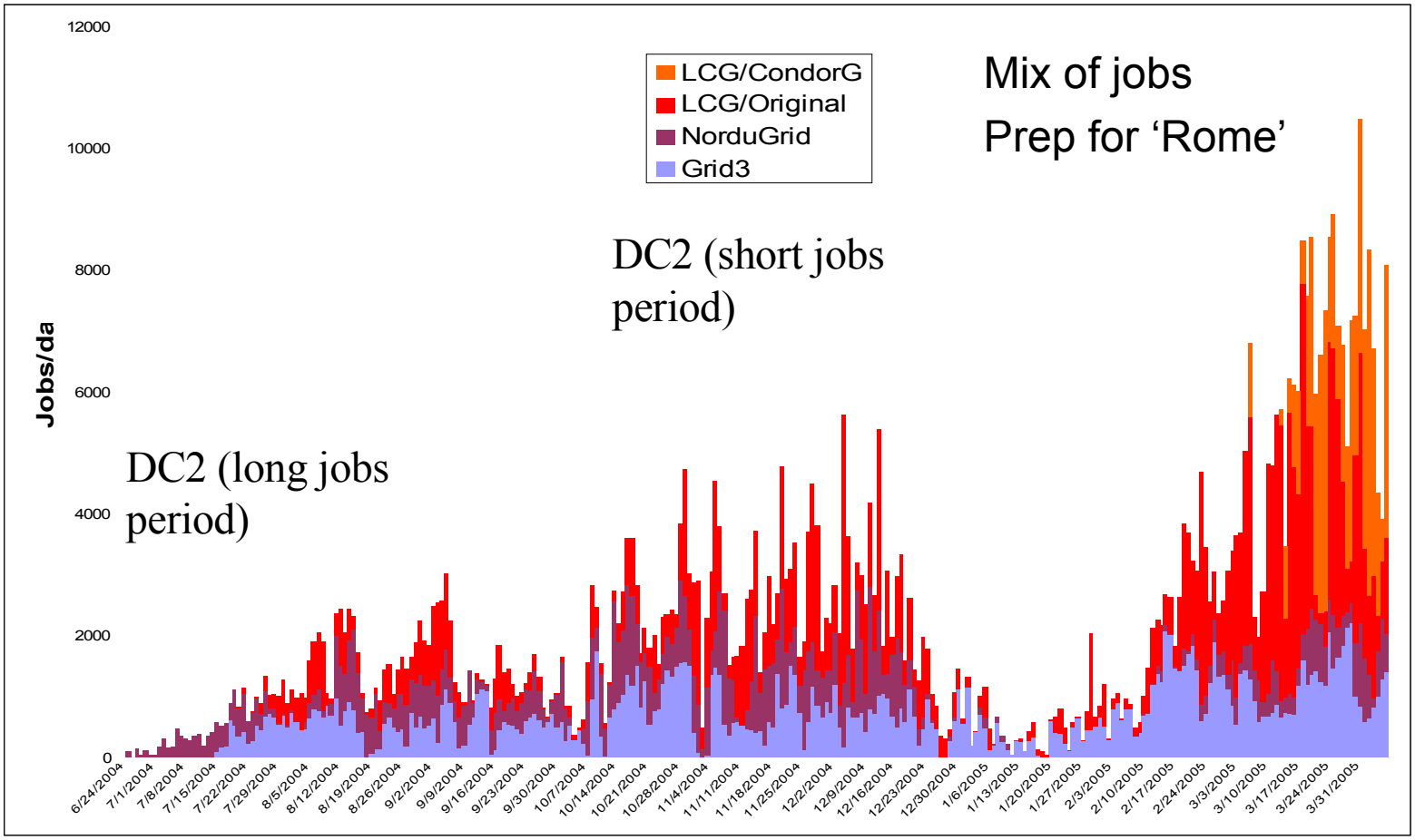
- Very experienced and large international user community
  - Involvement in many projects worldwide and users of several grids (e.g. all LHC experiments do use multiple grids at the same time for their data challenges)

Update on activities since Den Haag

- Production infrastructure (LCG/EGEE)
  - Intensive usage
    - E.g. LHCb 3500+ concurrent jobs for long periods (weeks), ATLAS 6000+ concurrent jobs (using multiple Grids as a backend: LCG, Grid3, NorduGrid), …
    - Complex simulation campaigns using high-level services developed in the HEP community especially for general steering, data distribution, monitor, usage of heterogeneous grids, …
    - Usage of the LCG2 infrastructure for analysis (system exposed to end users) has started (e.g. CMS)
    - Activity on several other HEP experiments using LCG2 (BaBar, CDF, D0, …)
- ARDA role in application development and middleware testing
  - Helping the evolution of the experiments specific middleware towards analysis usage
    - Large effort on the 4 LHC experiments' prototypes
    - CMS prototype migrated to gLite version 1 and exposed to several users
  - Early feedback on the utilisation of the gLite prototype right from the start of EGEE
  - Contribution to the common testing effort together with JRA1, SA1 and NA4-testing
    - Detailed performance/functionality measurements
    - Helping new colleagues to get up to speed (mini tutorial)

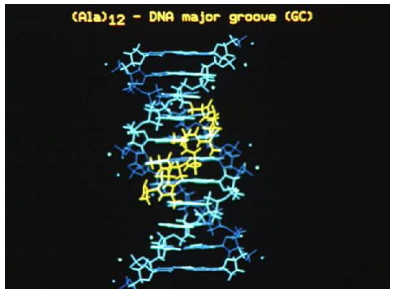
ATLAS grid work JUL 2004-MAR 2005

- ~ 660K jobs total in (LCG,Nordugrid,US Grid3)
- ~400 kSI2k years of CPU
- In latest period average ~7K jobs/day with ~5K in LCG

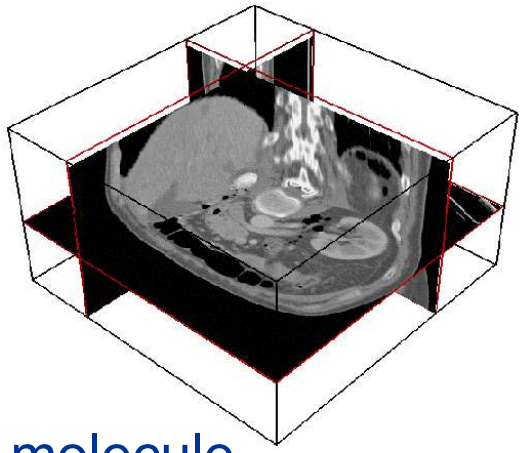


Biomed applications

- **Loosely coupled community**
- **Had to go the long way of getting up to speed**
  - VO creation and core services installation
  - Setting up a task force of experts
  - Recently joined the user support at application level



- **Applications**
  - See list and description from web site
    - http://egee-na4.ct.infn.it/biomed/applications.html
  - 12 applications running today



- **New applications emerging**
  - medical imaging, bioinformatics, phylogenetics, molecule structures and drug discovery...
- **Grown to a significant infrastructure usage**
  - 29kCPU hours and 24k jobs reported on January

Highlight on compute intensive biomed application

- **Drug discovery application**
  - LPC (CNRS, France) – SCAI (Fraunhofer Institute, Germany) collaboration
  - Targets malaria disease

- **Plans for data challenge (during summer)**

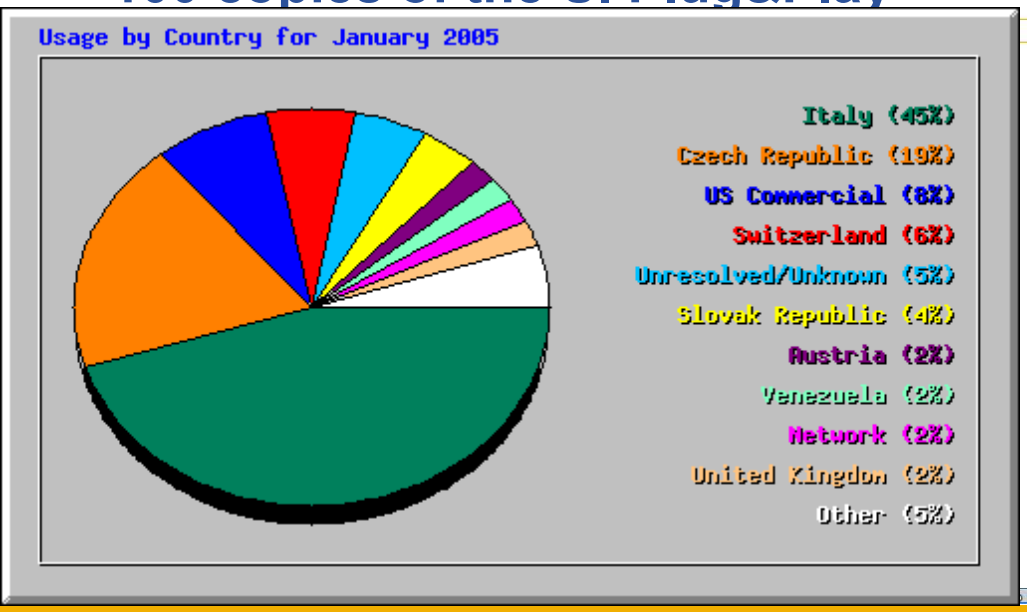
	Done	Challenge
Number of targets	1	5
Number of drug candidates	10^5	3 10^6
Total CPU time	188 days	80 years?
Gain of time	149	?

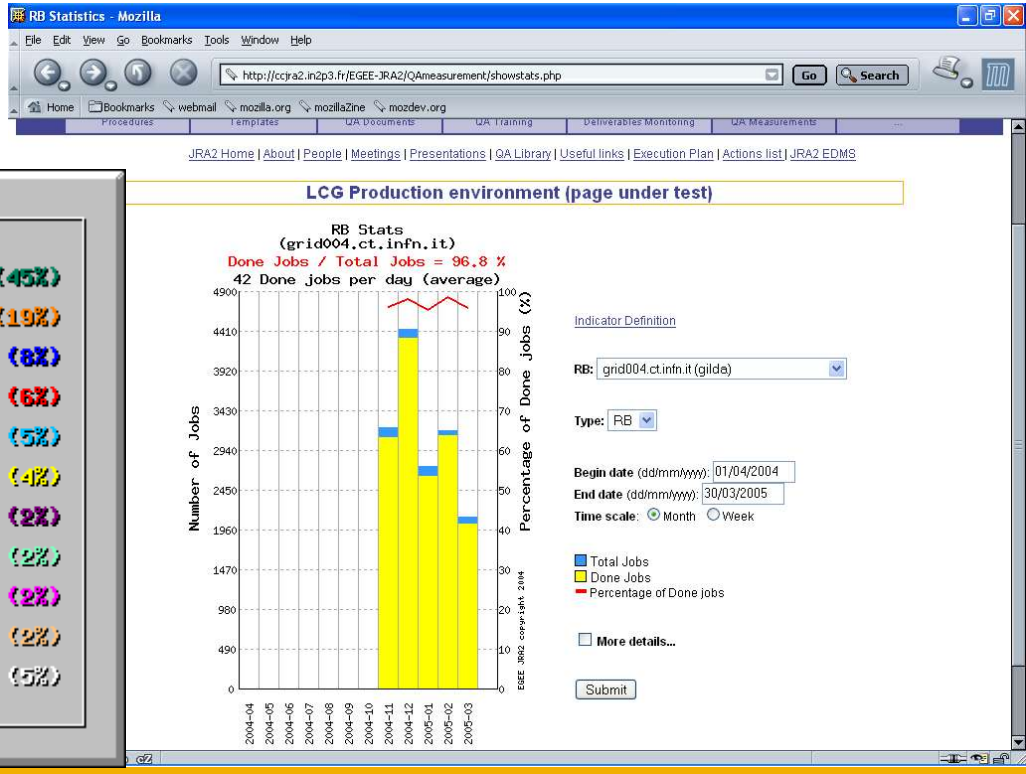
Generic applications

- **In fact many application areas**
  - Earth Observation, Geophysics, Astrophysics, Computational Chemistry, etc.
- **GILDA infrastructure**
  - https://gilda.ct.infn.it/
  - Lightweight users integration (certificates creation, resources provision...)
  - Handy catch-all infrastructure for first deployment and testing (no production)
  - gLite middleware recently installed (**demo** on wednesday)
  - 7 VO created
- **GENIUS web portal interface**
  - 3 biomed applications, 4 generic applications, 16 demonstrations
  - 1 complete CLI interface for NEMO
  - see https://grid-demo.ct.infn.it/

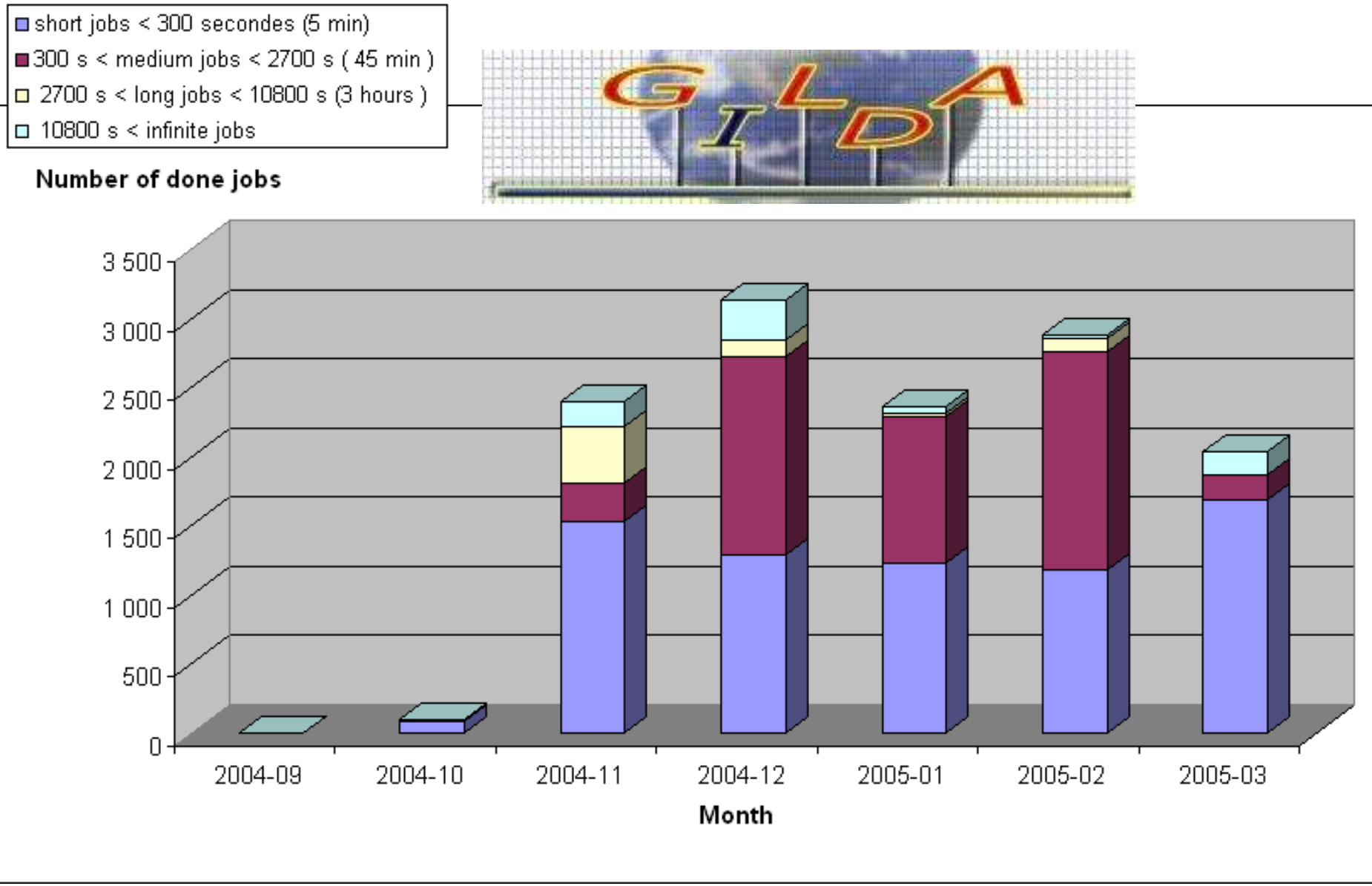
GILDA numbers (now)

- 14 sites in 2 continents
- >1400 certificates issued, 10% renewed at least once
- >40 tutorials and demos performed in 12 months
- >40 jobs/day on the average
- Job success rate above 96%
- >400,000 hits on the web site from 10's of different countries
- >200 copies of the UI live CD distributed in the world
- >100 copies of the UI Plug&Play





GILDA usage


short jobs < 300 secondes (5 min)
300 s < medium jobs < 2700 s (45 min)
2700 s < long jobs < 10800 s (3 hours)
10800 s < infinite jobs
Number of done jobs
3 500
3 000
2 500
2 000
1 500
1 000
500
0
2004-09
2004-10
2004-11
2004-12
2005-01
2005-02
2005-03
Month

NA4 organization

- **Application Working Group**
  - Management instance of NA4
  - Link to PEB
- **Project Technical Forum**
  - Technical instance including application representatives, middleware developers, and infrastructure providers
  - Application requirements collection http://savannah.cern.ch/support/?group=egeeptf
- **Infrastructure adoption and usage monitoring**
  - In collaboration with JRA2
  - Metrics definition and user surveys
- **Integration of new user communities**
  - OAG(Operations Advisory Group) joint group (talk by Rolf Rumler on Tuesday afternoon reporting on status and issues)

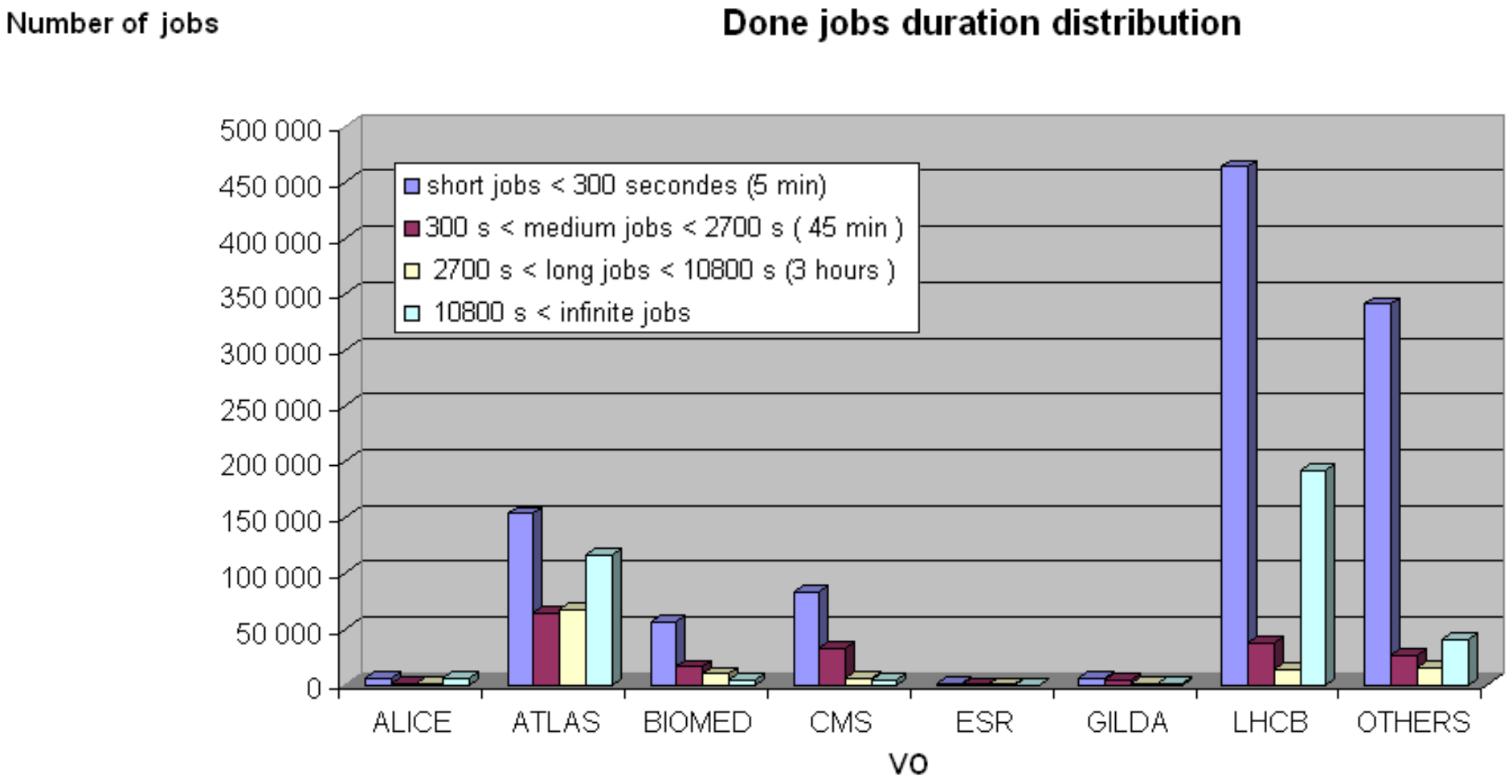
Memorandum of Understanding (MoU)

- **MoUs objectives**
  - identification of the main actors in the different organisations (EGEE NA4, NA3, SA1, SA2, Application…) in order to work in collaboration
  - Expression of application needs (CPU, storage, training…) and evaluation of their impacts on EGEE
  - EGEE and application commitments with deadlines on specific tasks
  - General evaluation of satisfaction and expectations
  - Document updated according to the evolution of the application
- **MoUs status**
  - MoUs being finalized with Magic, Planck, Comp Chemistry and Drug Discovery
  - Agreed MoUs for PM15 with EGEODE and Earth Sciences (academic and industry).
- **EGAAP meeting**
  - no call in next EGAAP
  - survey of applications status

LCG2 infrastructure usage

- **Job submission per VO**

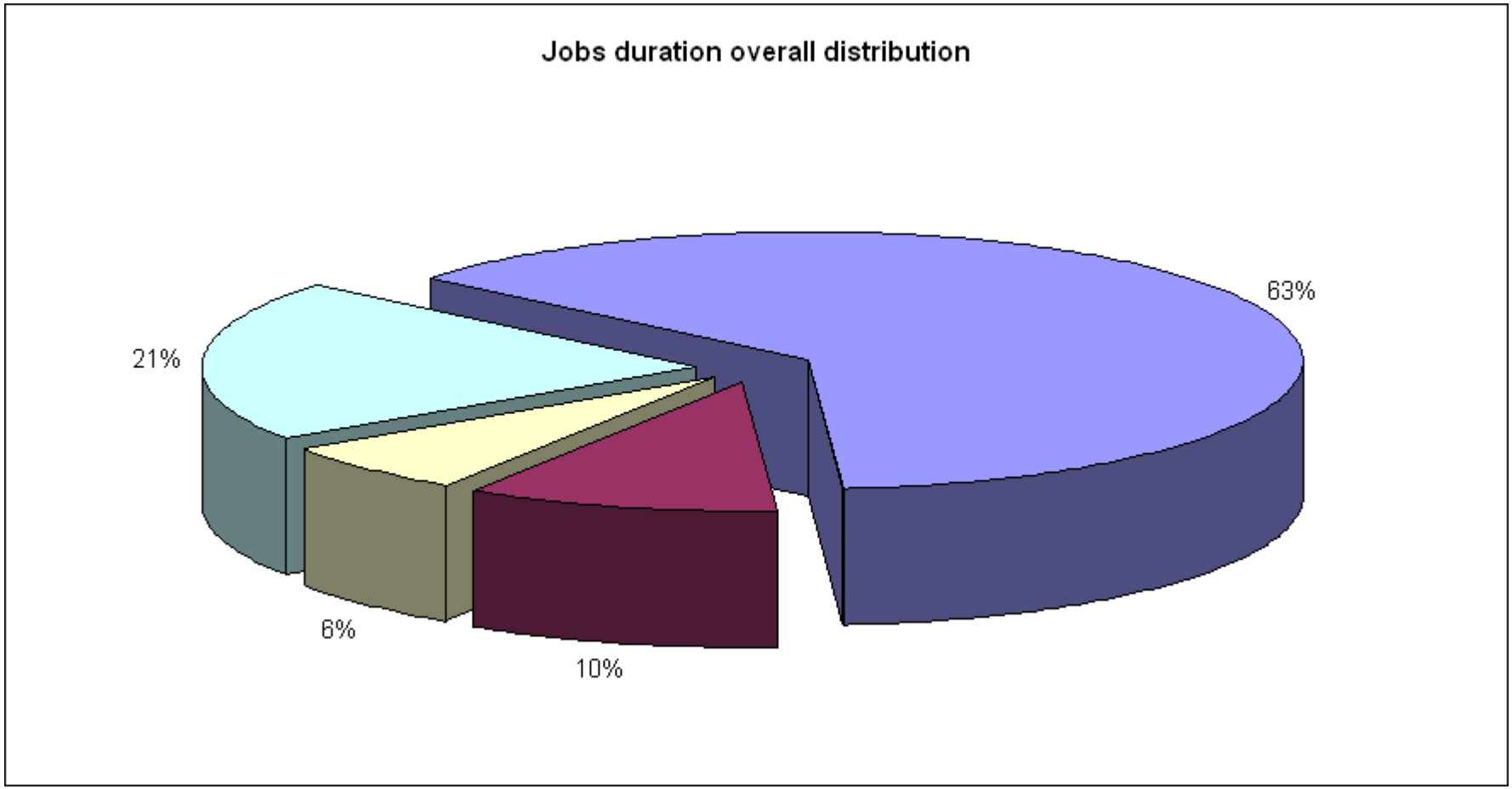
  http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/durationbyVO.php



LCG2 infrastructure usage

- **Job repartition per length**

  http://ccjra2.in2p3.fr/EGEE-JRA2/QAmeasurement/jobduration.php

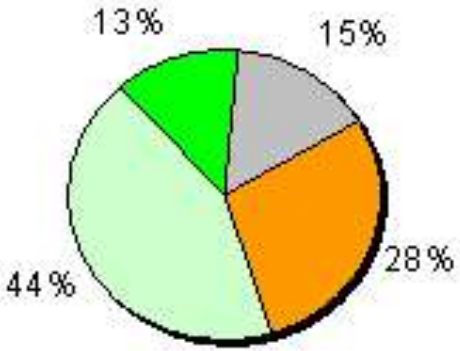


- **More statistics (VO per VO) available from JRA2 web**

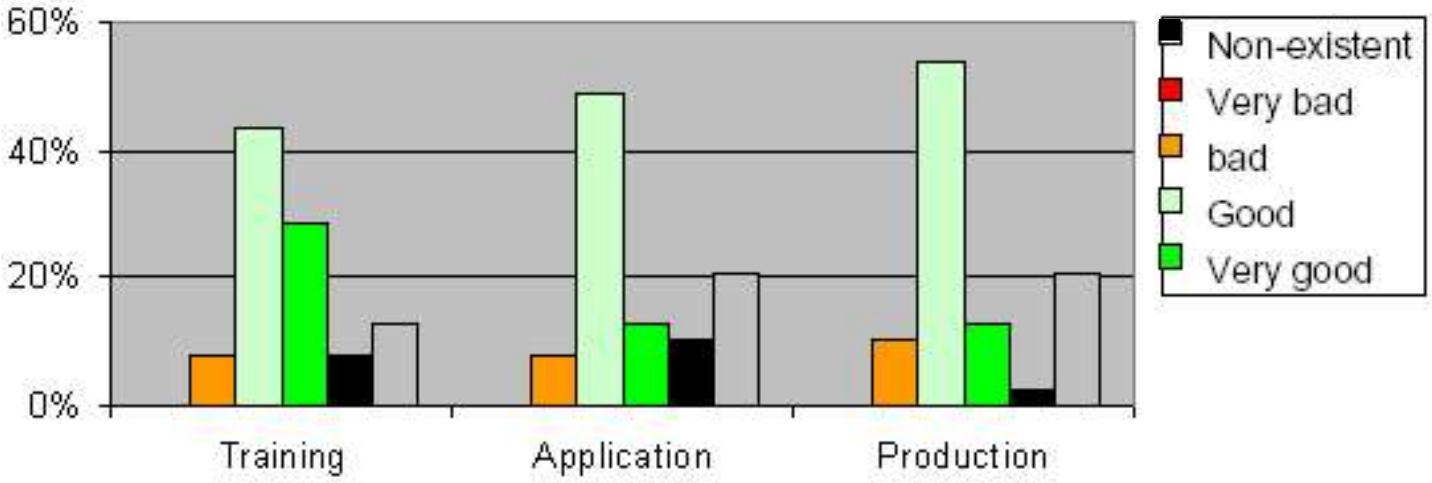
User feedback survey

- **User Survey (39 replies, 55% from biomed)**

Reliability



User Support



  - Shows user satisfaction level on support (very good), reliability of service (good), documentation (average)
- **User survey now available on line**
  - http://egee-na4.ct.infn.it/egee-survey-form/
  - more in Florence Jacq talk (NA2/3/4/5 session on thursday)

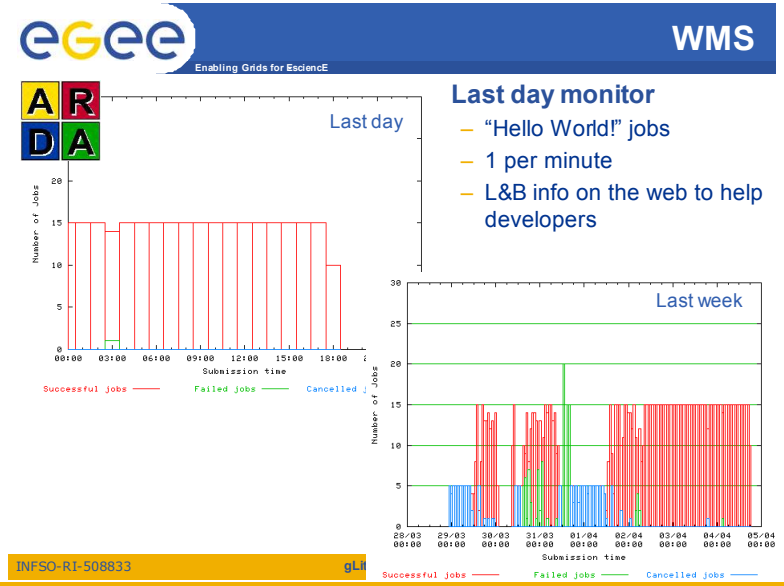
Highlights of the ARDA activity since Den Haag

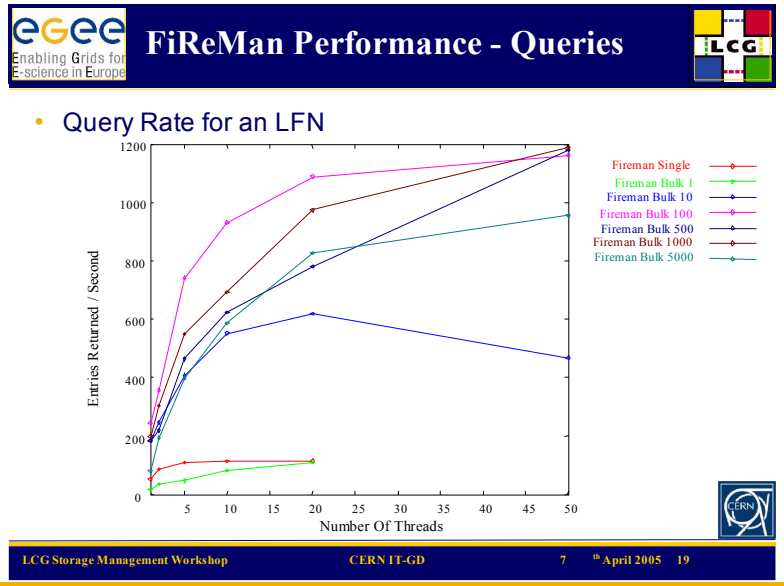
- More details in Massimo Lamanna's talk on Tuesday afternoon
- Info available also under http://lcg-web.cern.ch/lcg/PEB/arda/LCG_ARDA_Glite.html

- **gLite version 1**
  - WMS
    - *Continuous monitor available on the web (active since 17th of February)*
    - *Concurrency tests*
    - *Usage with ATLAS and CMS jobs (Using Storage Index)*
    - *Good improvements observed*
  - DMS (FiReMan + gLiteIO)
    - *Early usage and feedback (Starting Nov 2004) on functionality, performance and usability*
    - *Considerable improvement in performances/stability observed during the last months*
    - *Some of the tests given to the development team for tuning and to JRA1 to be used in the testing suite*
    - *Most of the tests given to JRA1 to be used in the testing suite*
    - *Performance/stability measurements: heavy-duty testing needed for real validation*
  - Contribution to the common testing effort to finalise gLite 1 (with SA1, JRA1 and NA4-testing)
    - *Migration of certification tests within the certification test suite (LCG→gLite)*
    - *Comparison between LFC (LCG) and FiReMan*
    - *Mini tutorial to facilitate the usage of gLite within the NA4 testing*

Experiment Prototypes

- **Experiment Prototypes activity**
  - Waiting for gLite 1.0 to be deployed on a larger infrastructure (pre production service)
  - Preparation work (migration) on the prototype (gLite 1.0 candidate)
- **HEP Highlights:**
  - Special effort on the CMS prototype. Several users involved.
  - Major redesign of Ganga (LHCb prototype). Large effort. Version 1.0 expected end of April





gLite usage and testing

- **gLite testing by biomedical applications**
  - Generic tests: difficulties to test in a development environment, need a pre-production stable infrastructure
  - VOMS, Data management, Installation...
  - Application deployment: will start on a limited number of applications
- **gLite testing by generic applications**
  - gLite successfully deployed on GILDA
  - first test jobs submitted
  - generic application soon able to port on gLite

Conclusion and plans for the coming year

- **Catching up with many user communities, many requirements**
  - HEP intensive testing
  - Biomed new VO integration
  - Generic community access eased by GILDA
- **Plans**
  - Migrating applications to gLite on the pre-production service
  - Validate 'satisfied' requirements
  - Deploying demonstrative applications
  - Supporting user communities on board
  - MoUs to implement virtuous cycle (PM15)
  - User survey completion (PM15)

... and beyond (EGEE 2)

- **Steer middleware development**
  - few selected pilots
- **Provide easy access to a wide variety of users**
  - emphasize on the GILDA role in EGEE
- **Define a high-level common application layer**
  - many high level requirements remain (application-level scheduling, interactivity, workflows, data integration...)
  - need for integrating other projects tools or application level development
- **Build multidisciplinary applications on top of high level services**
  - enable new applications to be developed
  - cross barriers between application areas